Ministry of Economic Development
Manatū Ōhanga
Intellectual Property Office of New Zealand

10/500158

#2

PCT/NZ03/00187

Rec'd PCT/PTO 25 JUN 2004

REC'D 22 SEP 2003
WIPO PCT

CERTIFICATE

This certificate is issued in support of an application for Patent registration in a country outside New Zealand pursuant to the Patents Act 1953 and the Regulations thereunder.

I hereby certify that annexed is a true copy of the Provisional Specification as filed on 23 August 2002 with an application for Letters Patent number 520986 made by THE UNIVERSITY OF WAIKATO.

Dated 10 September 2003.

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Neville Harris

Neville Harris
Commissioner of Patents, Trade Marks and Designs

COMMISSIONER OF PATENTS, TRADEMARKS & DESIGNS
NEW ZEALAND

330 High Street, Lower Hutt and 17 Toop Street, Seaview
PO Box 30687, Lower Hutt, New Zealand or DX SX 11129, Wellington
Phone: +64 4 560 1600 Fax: +64 4 560 1691, www.iponz.govt.nz

520906

PATENTS FORM NO. 4
Appln Fee: $50.00
James & Wells ref: 27764/16 CL

PATENTS ACT 1953
PROVISIONAL SPECIFICATION

AUDIOVISUAL MEDIA ENCODING SYSTEM

WE THE UNIVERSITY OF WAIKATO, a New Zealand company of Gate 5, Hillcrest Road, Hamilton, New Zealand.
do hereby declare this invention to be described in the following statement:

James & Wells Ref: 27764

AUDIOVISUAL MEDIA ENCODING SYSTEM

TECHNICAL FIELD

This invention relates to an Audiovisual Media Encoding System. Preferably, the present invention may be adapted to encode video conferences, seminars or presentations made over a computer network for review by an observer, either in real time or at a later time. Reference throughout this specification will also be made to the present invention being used in this situation, but those skilled in the art should appreciate that other applications are also envisioned and reference to the above only throughout this specification should in no way be seen as limiting.

BACKGROUND ART

Video conferencing systems have been developed which allow two-way audio and video communications between participants at remote locations. Participants may, through a common computer network, participate in a real time video conference with the assistance of cameras, microphones and appropriate hardware and software connected to the computer network used. Video conferences can be used to present seminars or other types of presentations where additional media such as slides or documents may also be supplied to a further input system or document camera for integration in the video signal sent.

As the participants of video conferences interact in real time with one another, this places a high demand on network bandwidth with the transmission of audiovisual content signals. Furthermore, there can be some quality problems with the audiovisual content of the conference if the network employed does not have the available bandwidth required to run the conference correctly. In such instances the internet protocol packets which make up the stream of signals between participants can be lost or late arriving to a recipient and hence cannot be integrated effectively in real time

into the video and audio played out.

In some instances it is also preferable to supply or stream these video conferencing signals to additional observers who do not necessarily participate in the conference. These observers may, for example, be interested in a seminar or presentation made but may not necessarily need to attend or participate in the conference in real time. Additional observers may view a stream of audiovisual signals in real time as the conference occurs, or alternatively can view this information at a later time as their participation within the conference is not required.

To stream video conference content to additional observers the signals generated are normally supplied to an additional encoding computer system. Using current technology such a computer is supplied with an analogue feed of the video and audio signals sourced from video conference unit cameras and microphones, which subsequently converts, encodes or formats this information into a digital computer system file which can be played by specific software player applications. The actual encoding or formatting applied will depend on the player application which is to subsequently play or display the encoded video conference. As can be appreciated by those skilled in the art, this encoded information may be streamed or transmitted out to observers in real time, or alternatively may be stored for later transmission to observers.

However, this approach used to encode video conference content for additional observers suffers from a number of problems.

In the first instance there are losses in accuracy or quality in the resulting formatted output due to the conversion of digital audio and video information to an analogue format for subsequent supply to the encoding computer system. In turn the computer system employed converts these analogue signals back into digital format, resulting in quality and accuracy losses with each conversion made.

Furthermore, the encoding computer used must be provided with an analogue cable connection to the video conferencing equipment and thereby in most instances must also be located within a room in which one end point of the video conference is to take place. This requires a further piece of apparatus to be located within the video conferencing room or suite, which must also be set up and configured prior to the conference in addition to the video conferencing equipment itself.

One attempt to address these issues has been made through use of a new video conferencing transmission protocol, being the H323 Version 4 umbrella protocol, as discussed in the ITU-T recommendation entitled "Packet-Based Multi-Media Communication System", published in November 2000. This protocol allows audiovisual signals and associated protocol information to be transmitted to a network address from the video conferencing equipment employed - without this network address acting as a full participant to the video conference call taking place. The additional connection can be described as a streaming end point for the video conference signals which can be supplied to the digital audio and visual information required, without the necessary digital to analogue to digital conversions required using existing technology.

However, this new protocol does not necessarily facilitate the encoding or re-formatting of audio and video signals in a readily usable form. A major complication with the use of this basic protocol arises from the high bandwidth requirements employed in the video conferencing call, and a subsequent streaming of signals to the end point at high bit rates. When re-transmitted to software player applications, the higher bit rate of the supplied input will be present in the output produced, thereby resulting in a large video file or high volume output, which cannot readily be accessed by low speed connections to the computer network employed.

An improved audiovisual media encoding system which addressed any or all of the above problems would be of advantage. A system would could act as an end point for

conference calls and could encode or format audio and video conference content for subsequent streaming or supply to observers would be of advantage. A system which could exhibit and provide flexibility and functionality regarding how these video and audio signals are encoded and supplied to observers would be of advantage.

All references, including any patents or patent applications cited in this specification are hereby incorporated by reference. No admission is made that any reference constitutes prior art. The discussion of the references states what their authors assert, and the applicants reserve the right to challenge the accuracy and pertinency of the cited documents. It will be clearly understood that, although a number of prior art publications are referred to herein, this reference does not constitute an admission that any of these documents form part of the common general knowledge in the art, in New Zealand or in any other country.

It is acknowledged that the term 'comprise' may, under varying jurisdictions, be attributed with either an exclusive or an inclusive meaning. For the purpose of this specification, and unless otherwise noted, the term 'comprise' shall have an inclusive meaning - i.e. that it will be taken to mean an inclusion of not only the listed components it directly references, but also other non-specified components or elements. This rationale will also be used when the term 'comprised' or 'comprising' is used in relation to one or more steps in a method or process.

It is an object of the present invention to address the foregoing problems or at least to provide the public with a useful choice.

Further aspects and advantages of the present invention will become apparent from the ensuing description which is given by way of example only.

DISCLOSURE OF INVENTION

According to one aspect of the present invention there is provided a method of

encoding audiovisual media signals, characterised by the steps of;

(i) receiving a video conference transmission from a computer network, said video conference transmission including an audiovisual signal or signals, and at least one protocol signal, and

(ii) reading one or more protocol signals, and

(iii) applying a selected encoding process to the received audiovisual signal or signals, said encoding process being selected depending on the contents of said at least one protocol signal read.

According to a further aspect of the present invention there is provided a method of encoding audiovisual media signals further characterised by the additional subsequent step of

(iv) producing encoded output for a software player application.

According to yet another aspect of the present invention there is provided a method of encoding audiovisual media signals substantially as described above, wherein the contents of said at least one read protocol signal indicates the time position or location of at least one key frame present within the audiovisual signal or signals of the video conference transmission.

According to a further aspect of the present invention there is provided a method of encoding audiovisual media signals substantially as described above, wherein the contents of said at least one read protocol signal indicates a content switch present within the audiovisual signal or signals of the video conference transmission.

According to a further aspect of the present invention there is provided a method of encoding audiovisual media signals substantially as described above, wherein the encoding process selected associates at least one index marker into said encoded

output when a content switch is indicated by said at least one read protocol signal.

According to a further aspect of the present invention there is provided a method of encoding audiovisual media signals substantially as described above, wherein the read protocol signal or signals provides information regarding any combination of the following parameters associated with the audiovisual signal or signals of the video conference transmission;

(i) audio codec employed and/or

(ii) video codec employed and/or

(iii) the bit rate of audio information supplied and/or

(iv) the bit rate of video information supplied and/or

(v) the video information frame rate and/or

(vi) the video information resolution.

The present invention is preferably adapted to provide a system and method for encoding audiovisual media signals. Preferably these signals may be sourced or supplied from a video conference transmission, with the present invention being adapted to encode at least a portion of these signals into a format which can be played to other users or observers who are not directly participating in the video conference.

Preferably, the present invention may be adapted to provide an encoded output file, signal or transmission, which can be received or played by a computer based software player application to display audiovisual media or content. The encoded output provided using the present invention may, in some instances be streamed or transmitted to non-participating observers of a video conference in real time as the video conference occurs. Alternatively, in other instances, the encoded output provided may be saved to a computer file which in turn can be downloaded or

transmitted to non-participating observers to be played at a later time.

For example, in some instances the present invention may be adapted to provide an encoded audiovisual content output which can be played with Microsoft's Windows Media Player ™, Apple's Quicktime Player ™, or Real Network's RealPlayer ™. Furthermore, the players involved may also support the real time streaming of the encoded output to observers as the video conference involved occurs.

Reference throughout this specification will also be made to the encoded output provided being adapted to provide an input for a software based player application for a computer system. However, those skilled in the art should appreciate that other formats or forms of encoded output may also be produced in conjunction with the present invention and reference to the above only throughout this specification should in no way be seen as limiting. For example, in other embodiments the present invention may provide an encoded output which can be played using a cellular phone, PDA's, game consoles or other similar types of equipment.

Preferably, the video conference transmissions made may be transmitted through use of a computer network. Computer networks are well-known in the art and can take advantage of existing transmission protocols such as TCP/IP to deliver packets of information to participants in the video conference.

In a preferred embodiment, the video conference transmissions received in conjunction with the present invention may be supplied through a computer network as discussed above. Receiving and encoding hardware employed in conjunction with the present invention may be connected to such a computer network and assigned a particular network or IP address to which these video conference transmissions may be delivered.

Those skilled in the art should appreciate that reference to computer networks throughout this specification may encompass both networks provided through dedicated ethernet cabling, wireless radio networks, and also distributed networks

which employ telecommunications systems.

In a further preferred embodiment, hardware or apparatus employed by the present invention may be described as a streaming or streamed end point for the video conference call involved. A streaming end point may act as a participant to the video conference without necessarily supplying any usable content to the video conference call. This end point of a particular address in the computer network may therefore receive all the transmissions associated with a particular video conference without necessarily contributing usable content to the conference.

The present invention preferably provides both a method and apparatus or system for encoding audiovisual media. The system or apparatus employed may be formed from or constitute a computer system loaded with (and adapted to execute) appropriate encoding software. Such software (through execution on the computer system through the computer system's connections to a computer network) can implement the method of encoding discussed with respect to the present invention. Furthermore, this computer system may also be adapted to store computer files generated as an encoded output of the method described, or retransmit the encoded output provided to further observers in real time.

Reference throughout this specification will also be made to the present invention employing or encompassing an encoding computer system connected to a computer network which is adapted to receive video conference transmissions and to encode same using appropriate software.

For example, in one instance the present invention may take advantage of the H323 protocol for video conference transmissions made over a computer network. This protocol may be used to supply digital signals directly to an encoding computer system without any digital to analogue to digital conversions of signals required.

Reference throughout this specification will also be made to the present invention

being used to encode audiovisual media sourced from a video conference transmission made over a computer network. However, those skilled in the art should appreciate that other applications are envisioned for the present invention and reference to the above only throughout this specification should in no way be seen as limiting. For example, the present invention may be used to encode other forms of streamed or real time audiovisual transmissions which need not necessarily be video conference based, nor directly related to transmissions over computer networks.

Preferably, the video conference transmissions received by the encoding computer may be composed of or include at least one audiovisual signal or signals and at least one protocol signal or signals.

Preferably, an audio visual signal may carry information relating to audio and/or video content of a video conference as it occurs in real time. A single signal may be provided which carries both the audio and visual content of the conference as it is played out over time in some instances. However, in alternative situations a separate signal may be provided for both the audio and the video components of such conferences required.

Preferably, the video conference transmissions received also incorporates or includes at least one protocol signal or signals. A protocol signal may carry information relating to the formatting or make up of an audiovisual signal, including parameters associated with how such a signal was generated, as well as information relating to the configuration, status, or state of the physical hardware used to generate such a signal. Furthermore, a protocol signal may also provide indications with regard to when the content displayed changes or switches using feedback or information from the particular hardware used to generate an audiovisual signal. In addition, a protocol signal may also provide information regarding how a transmitted audiovisual signal was created such as, for example, whether a data compression scheme was used in the generation of the signal, and also may provide some basic information regarding how

such a compression scheme operated.

Preferably, the present invention may be adapted to initially read at least one protocol signal received in conjunction with an audiovisual signal making up the video conference transmission. The particular information encoded into such a protocol signal or signals can then be used to make specific decisions or determinations regarding how the incoming audiovisual signal should in turn be encoded or formatted for supply to further observers. The information harvested from a protocol signal can be used to select and subsequently apply a specific encoding process or algorithm to produce the encoded output required of the present invention. The exact form of the information obtained from the protocol signal and the encoding processes available and of interest to an operator of the present invention will determine which encoding process is selected and applied.

In a preferred embodiment, information obtained from a protocol signal may include or indicate the time position or location of key frames present within the audiovisual signal or signals received.

Key frames are generated and used in digital video compression processes, and provide the equivalent of a full traditional video frame of information. In addition to key frames, pixel modification instructions can be transmitted as the second portion of the video information involved. A key frame (which incorporates a significant amount of data) can be taken and then further information regarding the change in position of objects within the original key frame can then be sent over time, thereby reducing the amount of data which needs to be transmitted as part of an audiovisual signal.

This approach to video compression does however approximate the actual frames which composed the original video signal, as whole original frames (the key frames) are only transmitted or incorporated occasionally. If a previously compressed video signal is subsequently re-encoded or re-formatted these key frames may be lost or a

new key frame may be selected which was not originally a key frame in the starting compressed video. This can degrade the quality or accuracy of the resulting re-encoded or re-formatted video signal. However, if in conjunction with the present invention, the time position or locations of each of the key frames employed can be extracted from protocol information. This allows the same key frames to then be re-used in the re-encoding or re-formatting of the video content of the audiovisual signal while minimizing any subsequent loss of quality or introduction of further inaccuracies.

In a preferred embodiment, the present invention may also provide the user interface facility which allows a user or operator to set up how they would prefer incoming audiovisual signals to be encoded or formatted. An operator may supply required parameters or input information with such a user interface, which can in turn be used to tailor the characteristics of the encoded output produced.

In a further preferred embodiment, information or parameters regarding the characteristics of an incoming audiovisual signal may also be extracted from one or more protocol signals. This information may be used in conjunction with information supplied by a user to determine a potential encoding scheme or schemes to be selected in a particular instance. In yet a further preferred embodiment, the present invention may include the facility to pre-calculate or pre-assess a number of encoding schemes which will potentially produce the best resulting encoded output based on both information received from a user and obtained from a protocol signal or signals.

This facility can in fact operate like a user interface "wizard" so that the user will be presented with a facility to select and use only encoding schemes which are capable of satisfying the user requirements or parameters supplied based on the information extracted from a protocol signal or signals associated with an incoming video conference transmission.

For example, in one preferred embodiment, a user may input a required bit rate for the resulting encoded output in addition to the software player format required for the resulting output. Further information may also be provided by a user with respect to the number of monitors they wish to simulate from the video conference call.

Information regarding the make-up or characteristics of an incoming audiovisual signal can then be obtained from one or more protocol signal or signals. For example, in one instance, this information obtained from a protocol signal may include any combination of the following;

(i) audio codec employed

(ii) video codec employed

(iii) audio bit rate

(iv) video bit rate

(v) video frame rate

(vi) video resolution.

This information available for the software associated with or used by the present invention can then make a selection or present a range of options to a user indicating which audio and/or video codec to use, as well as the particular video resolution and video frame rates available for use which will satisfy the input criteria originally supplied by the user.

In a preferred embodiment information may be obtained from at least one protocol signal which indicates a content switch present within the audiovisual signal or signals received. Such a content switch may indicate that audiovisual signals are generated by a new or different piece of hardware, or that the configuration of a currently used camera or microphone has been modified.

For example, in some instances a protocol signal may indicate that a freeze picture control signal has been sent as part of a video conference transmission. This freeze signal will hold the current frame or picture making up the video content of the conference on the screens of all participants. The transmission of an un-freeze controls command within a protocol signal may also be detected as a content switch in conjunction with the present invention.

Furthermore, a content switch may also be detected through a protocol signal indicating whether a document camera is currently being used to provide a video feed into the conference. Such a document camera may show good quality close views of printed material as opposed to the participants of the conference. As such, the activation or use of a document camera will in turn indicate that the content of the video signals transmitted has switched or changed.

In yet another instance a protocol signal may carry status information indicating that a digital image or digital slide is to be used to currently form the video content of the conference. Again, a still image or 'snap shot' may be presented as the video content of the conference with this image sourced from a digital file or source - as opposed to a document camera as discussed above.

Furthermore, content switches may also be detected through the automated panning or movement of a video camera lens from a number of pre-selected viewing positions or angles. These viewing positions may be pre-set to focus a camera on selected seating positions and their associated speakers, so that when the camera preset viewing angle changes, the content switch involved can be indicated by information present within a protocol signal.

In a further preferred embodiment, the detection or indication of a content switch within an audiovisual signal may trigger the association of at least one index marker with the encoded output provided, where this index marker is associated with

approximately the same time position in the encoded output as the content switch was detected in the incoming audiovisual signal or signals.

Furthermore, the particular index marker encoded may also include reference information regarding how the particular content switch was detected and therefore may give an indication as to what the content of the audiovisual signal is at the particular time position which the index marker is located at.

In a preferred embodiment an index marker may be associated with the encoded output provided through the actual encoding of a reference, pointer or other similar marker actually within the encoded output provided. This marker or reference may then be detected by a player application at approximately the same position as the content switch of the video content in place. However, in other embodiments an index marker may not necessarily be directly encoded into the output to be provided. For example, in one embodiment a log file or separate record of index markers may be recorded in addition to time position or location information associated with the video signal involved. This file can indicate at which particular time positions an index marker is associated with the video content involved.

In a further preferred embodiment, an index marker may be implemented through the insertion of a universal resource locater (URL) into the encoded output produced by the present invention. Those skilled in the art should appreciate that URL's are commonly used in the art to index audiovisual media, and as such the present invention may employ existing technology to implement the index markers discussed above.

Preferably, these index markers encoded into the output provided may be used by the user of a player application to proactively seek or search through the audiovisual output of the present invention, depending on the particular content which these index markers reference. An index marker may mark the time position or location in the encoded output at which selected types of content are present and subsequently allow a

user to easily search the entire output produced for a selected portion or type of content.

In a further preferred embodiment, the presence of original key frames within an incoming audiovisual signal or signal's in proximity to the time position at which an index marker is to be encoded can also be detected in conjunction with the present invention.

If too many key frames are located in proximity to one another this will degrade the quality of resulting coded output of the present invention, and also potentially increase its size or volume. However, it is preferable to have a key frame close to an index marker in the encoded output as this will allow a software player application to seek to the time position of the index marker to quickly generate the video content required using a nearby key frame.

Preferably, through detecting whether an original key frame is near to the time position at which an index marker is to be encoded, the present invention may optimise the placement of key frames in the resulting encoded output. If no key frame is present within a specified time displacement tolerance, a new key frame may be encoded at approximately just before, or at the same time position as where the index marker is to be encoded. Conversely, if an appropriate nearby key frame is available, no new key frame may be generated or incorporated into the resulting encoded output.

In a preferred embodiment, the present invention may also be used to modify the timing or time position of particular portions of audiovisual content present within the encoded output when compared to the original audiovisual signal or signals provided. This timing modification may be completed if a particular content switch is detected through reading a protocol signal or signals.

In a further preferred embodiment the video and audio content received may be time compressed if a freeze or hold picture control instruction is detected in a protocol

signal. Normally hold or freeze picture instructions are associated with the transmission of large amounts of image information between participants in the video conference, which can take some time to arrive and be assembled at a particular end point. This in turn can provide a relatively stilted presentation as the participant's interest in the current frozen image or picture may have been exhausted before all of this information has been received and subsequently displayed. Conversely, through use of a data buffering system employed in conjunction with the present invention, this information system may be pre-cached and subsequently displayed for a short period of time only. The audio content of the conference may also be compressed over time to synchronise the audio and visual content portions, provided that limited audio content is also generated over the time at which the still image or frozen frame is displayed.

The present invention may provide many potential advantages over the prior art.

The present invention may read and subsequently employ information from a protocol signal or signals to make intelligent decisions regarding how an audiovisual signal or stream should be encoded or re-formatted.

Information may be obtained from such protocol signals regarding the original key frame placement within the incoming audiovisual signal, with this information in turn being employed to re-use the same key frames in output audiovisual information provided. Furthermore, this technique may also be of assistance where particular content switches within the received audiovisual signal are detected and indexed in the encoded output provided. These index markers supplied can allow a user to proactively seek or search through the resulting encoded output quickly for particular types of content. Furthermore, the key frame placement information obtained from a protocol signal can also be used to ensure that a key frame is placed in close time proximity to such index markers, thereby allowing the video information required to be generated and displayed quickly to a user.

Information obtained from a protocol signal or signals may also be used to assist in the selection of a particular encoding scheme or profile for an incoming audiovisual signal or signals. Based on user preferences or selections and in conjunction with information relating to the characteristics of an incoming audiovisual signal obtained from a protocol signal, a user may be presented with a limited number of coding schemes which will produce the best results for the input information that is supplied.

The present invention may also provide a facility to compress with respect to presentation time selected types of content present with an incoming audiovisual signal or signals. If a relatively stilted or slow content portion is detected within an incoming video conference (such as a freeze picture segment) the time over which the content is present may be compressed in the encoded output provided.

BRIEF DESCRIPTION OF DRAWINGS

Further aspects of the present invention will become apparent from the following description which is given by way of example only and with reference to the accompanying drawings in which:

Figure 1 shows a block schematic flowchart diagram of steps executed in a method of encoding audiovisual media signals in conjunction with a preferred embodiment, and

Figure 2 illustrates in schematic form signals involved with the encoding process discussed with respect to Figure 1, and

Figures 3a, 3b, 3c show in schematic form signals with encoded key frames as discussed with respect to Figure 2.

Figure 4 shows a user interface and encoding scheme selection facility provided in accordance with another embodiment of the present invention.

Figures 5a, 5b, 5c show a series of schematic diagrams of signals both used and produced in accordance with a further embodiment of the present invention, and

Figures 6a & 6b again show schematically a set of signals received and subsequently produced in accordance with yet another embodiment of the present invention.

BEST MODES FOR CARRYING OUT THE INVENTION

Figure 1 shows a block schematic flowchart diagram of steps executed in a method of encoding audiovisual media signals in conjunction with a preferred embodiment.

In the first step of this method an encoding computer system connected to a computer network receives a video conference transmission from the computer network. This video conference transmission includes audiovisual signals and a set of protocol signals. The protocol signals provide information regarding how the audiovisual signals were generated, in addition to the status of the particular hardware equipment used to generate signals.

In stage two of this method, information is extracted from the protocol signals received in stage 1. In the embodiment discussed with respect to Figures 1 and 2, the information extracted from these protocol signals consists of an indication of the time position at which key frames are encoded into the original audiovisual signals received and also information regarding when a particular content switch occurs within the audiovisual information employed. In the embodiment considered a content switch is detected through the use of a document camera as opposed to a camera which shows the participants of the conference.

At stage three of this method a specific encoding process is selected for application to the received audiovisual signals based on the information present within the protocol

signals read. In the instance discussed, the encoding process selected incorporates specific index marker references into the output provided to indicate the content switch present within the audiovisual information when a document camera is used. The encoding process selected also takes into account the position of each of the key frames encoded into the original audiovisual signal and adjusts its generation or application of key frames within the encoded output produced based on the time positions of the original key frames used.

In step four of this method the encoded output of the method is generated and produced for a particular software player application. In the instance discussed with respect to Figures 1 and 2, encoded output provided may be played on a Real Media Real Player.

Figure 2 illustrates in schematic form elements of the encoding process discussed with respect to Figure 1, showing an original audiovisual signal 5 and subsequent encoded output audiovisual signal 6.

The original signal 5 includes a number of key frames 7 distributed at specific time positions along the playing time of the signal 5. The original signal 5 also incorporates specific content switches between a video showing content participants 8 and a still image or snap shot 9 taken from the video camera trained on the conference participants.

The re-encoded signal 6 takes advantage of information obtained from protocol signals received from an incoming video conference transmission to detect the presence of the key frames 7 and content switches taking place. Index markers 10 (formed in a preferred embodiment by URL's) are inserted into the encoded output signal 6 to indicate the presence of a content switch in the audiovisual content of the signal.

Where possible, the original key frames 7 of the incoming audiovisual signal 5 are also recycled or reused as shown by the placement of the first key frame 11a in the second

signal 6. However, in the instance shown, a new key frame 11b is generated and encoded into the second signal 6 to provide a key frame in close proximity to an index marker indicating the presence of a content switch in the audiovisual information to be displayed. In this instance the second key frame 7b of the original signal is not re-encoded or reused within the second signal 6.

Figures 3a through 3c show an incoming video stream (3a), a video stream which is re-encoded without use of the present invention (3b) and a video stream re-encoded using the present invention (3c) where information regarding the original key frame placements of the original video stream (3a) is employed.

As can be seen from Figure 3b, a transcoded or re-encoded video signal does not necessarily have key frames placed at the same positions or locations as those provided in the signal shown with respect to Figure 3a without use of the present invention. Conversely, in Figure 3c key frames employed are positioned at essentially the same time position as the original key frames within the original streamed video signal.

Figure 4 shows a user interface and encoding scheme selection facility provided in accordance with another embodiment of the present invention.

In the instance shown an encoding computer system 12 is provided with a connection 13 to a computer network 14. This computer network 14 can carry video conference transmissions to be supplied to the encoding computer 12 which acts as an encoding end point for the video conference. However, the encoding computer 12 does not transmit any video or audio signals as a participant to the conference, and is adapted to provide further encoded audiovisual output sourced from the audiovisual signals employed within the video conference transmission.

A user interface module 15 may be provided in communication with the encoding computer 12 for a separate user computer, or through software running on the same encoding computer 12. This user interface (UI) module can initially send user

parameter information 16 to the encoding computer system. The encoding computer system 12 can also extract audiovisual signal parameter information from protocol signals received as part of the video conference transmissions, where these parameters give information regarding the audiovisual signals making up part of the video transmission. These parameters can provide information relating to the make up of an incoming audiovisual signal such as;

(i) the audio codec employed, and

(ii) the video codec employed, and

(iii) the bit rate of audio information supplied, and

(iv) the bit rate of video information supplied, and

(v) the video information frame rate, and

(vi) the video information resolution.

The encoding computer system may, using all of the user and protocol information obtained, calculate a number of "best fit" encoding schemes which can be used to meet the requirements of a user for an incoming video stream. Information regarding valid encoding schemes may then be transmitted 17 to the UI module, which in turn allows a user to transmit the scheme selection instruction 18 back to the encoding computer 12 to indicate which encoding scheme should be employed.

Based on these instructions, the encoding computer system may encode and produce output 19 which can be played on a suitable computer based media player application.

The process used to select or specify a set of encoding schemes which may be used is also shown in more detail through the pseudo code set out below.

```
H.323 call parameters:
    H.263 video @ 112kbps
    H.263 video resolution @ CIF
```

H.263 video frame rate @ 12.5fps
G.728 audio @ 16kbps

User input:

Bitrate: 56kbps Modem
Player format: RealMedia Native - Single Stream
Display mode: Single Monitor

Profiler decisions:

```
// find the media type for the stream
// either standard (video and audio only) or presentation (audio, video and
// snapshots)
If Display_Mode = Single_Monitor then
        Profiler_Media_Type = (standard)
Else
        Profiler_Media_Type = (presentation)
EndIf

// find the maximum audio bitrate for the stream based on the media type
// where media type is standard, allow more bitrate to the audio codec than if
// media type of presentation selected (when presentation need to leave
// bandwidth for the snapshot).
User_Bitrate = (56kbps) and Profiler_Media_Type = (standard) therefore
Max_Audio_Bitrate = (8.5kbps).

// select the audio codec for use in the stream based on the maximum
// available bandwidth.
If Incoming_Audio_Bitrate > Max_Audio_Bitrate then
    Profiler_Audio_Codec = Select Audio_Codec from Table_3 where
    Bitrate_Supported <= Max_Audio_Bitrate therefore
    Profiler_Audio_Codec = (RealAudio_8.5kbps_Voice)
Else
    Profiler_Audio_Codec = Incoming_Audio_Codec
EndIf

// set the video bandwidth based on total available bandwidth and bandwidth
// used by audio codec.
Profiler_Optimum_Bitrate = Select Optimum_Bitrate from Table_4 where
Bandwidth_Option = (56kbps_Modem)

If (Profiler_Audio_Codec <> Incoming_Audio_Codec) then
    Profiler_Audio_Bitrate = Select Bitrate_Supported from Table_3 where
    Audio_Codec = (Profiler_Audio_Codec)
Else
    Profiler_Audio_Bitrate = Incoming_Audio_Bitrate
EndIf

Profiler_Video_Bitrate = Profiler_Optimum_Bitrate - Profiler_Audio_Bitrate
therefore
Profiler_Video_Bitrate = (29.5kbps)

// set video resolution
Profiler_Video_Res = Select Optimum_Resolution from Table_4 where
Bandwidth_Option = (56kbps_Modem) therefore
Profiler_Video_Res = (176x144)
```

```
// set video codec
If User_Player_Format = RealMedia_Native then Profiler_Video_Codec =
(RealVideo9).

// set video frame rate
Max_Profiler_Frame_Rate = Incoming_Frame_Rate
Profiler_Frame_Rate = Select Optimum_Frame_Rate from
Table_4 where Bandwidth_Option = (56kbpsModem)
If Profiler_Frame_Rate > Max_Profiler_Frame_Rate then
        Profiler_Frame_Rate = Max_Profiler_Frame_Rate
EndIf
```

Figures 5a through 5c show a series of schematic diagrams of signals associated with the present invention, and illustrate further behaviour of the invention depending on the input signals it receives.

Figure 5a shows an incoming protocol signal which indicates that a snap shot event is to occur at frame 150 of the video signal shown with respect to Figure 5b. Figure 5b also shows that a key frame has been encoded into the original incoming video at frame 125.

Figure 5c shows the encoded video output provided in conjunction with the present invention in the embodiment shown. This figure illustrates how the invention can be used to place a key frame in its encoded output signal depending on the input the video conference transmissions received.

The software employed by the present invention makes a set of decisions in the instance shown. The first of these decisions is completed through considering a set value for the maximum time displacement between key frames which should be in the encoded output signal. In the instance shown a key frame is to be encoded every one hundred and fifty frames, and as a key frame is provided at frame 124, this original key frame is subsequently used in the encoded output 5c.

Secondly, the software employed notes that an index marker is to be encoded or written to the output provided at frame 150 to mark the placement of the snap shot event in the incoming video signal. By considering a tolerance value for time

displacement from this index marker, the software employed can see that the key frame present at frame 124 is within this tolerance and an additional key frame does not necessarily need to be encoded just before the snap shot event at frame 150.

Figures 6a and 6b show a set of signals illustrating further behaviour of the present invention in yet another embodiment. In the embodiment shown an incoming video signal is shown with respect to Figure 6a, whereas the encoded output video provided in conjunction with the present invention is shown as Figure 6b.

The incoming video begins at frame 248 and includes a pair of key frames around frames 275 and 405. Conversely, the encoded output provided includes key frames at frame 248 and frame 405 respectively. In the instance shown a decision is made to encode the output to be provided so that key frames are located at approximately every 150 frames. However, this maximum time between key frames may be varied depending on the particulars of the incoming signal, as discussed below.

When the original key frame located at frame position 275 in the incoming signal is detected, a decision is made by the software employed not to encode a key frame in the output due to the proximity to the first encoded key frame provided at frame 248. One hundred and fifty frames from frame 248, a key frame is not immediately encoded in the output as the software employed knows from the incoming video stream that key frames are encoded in this signal every one hundred and thirty eight frames, and therefore a further key frame will be supplied at frame 405. In this instance the maximum time between key frames is extended slightly to allow the original key frame to be employed again in the encoded output provided.

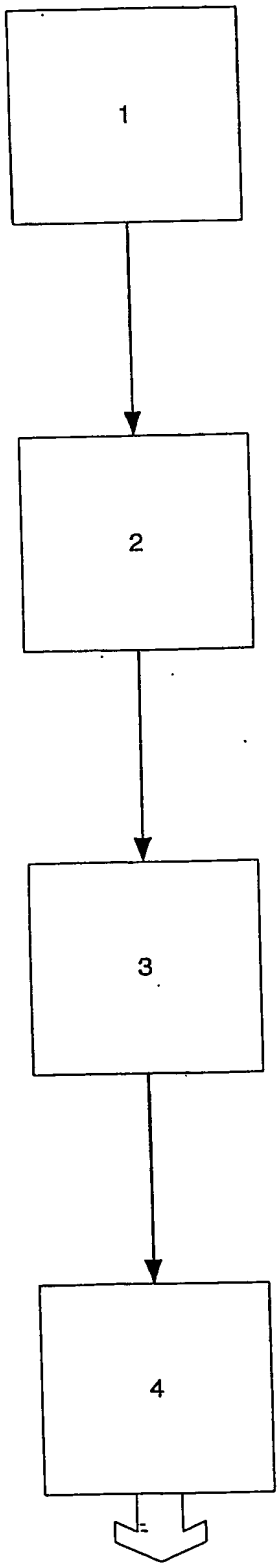
Aspects of the present invention have been described by way of example only and it should be appreciated that modifications and additions may be made thereto without departing from the scope thereof.

THE UNIVERSITY OF WAIKATO

by their Attorneys

JAMES & WELLS

Fig 1



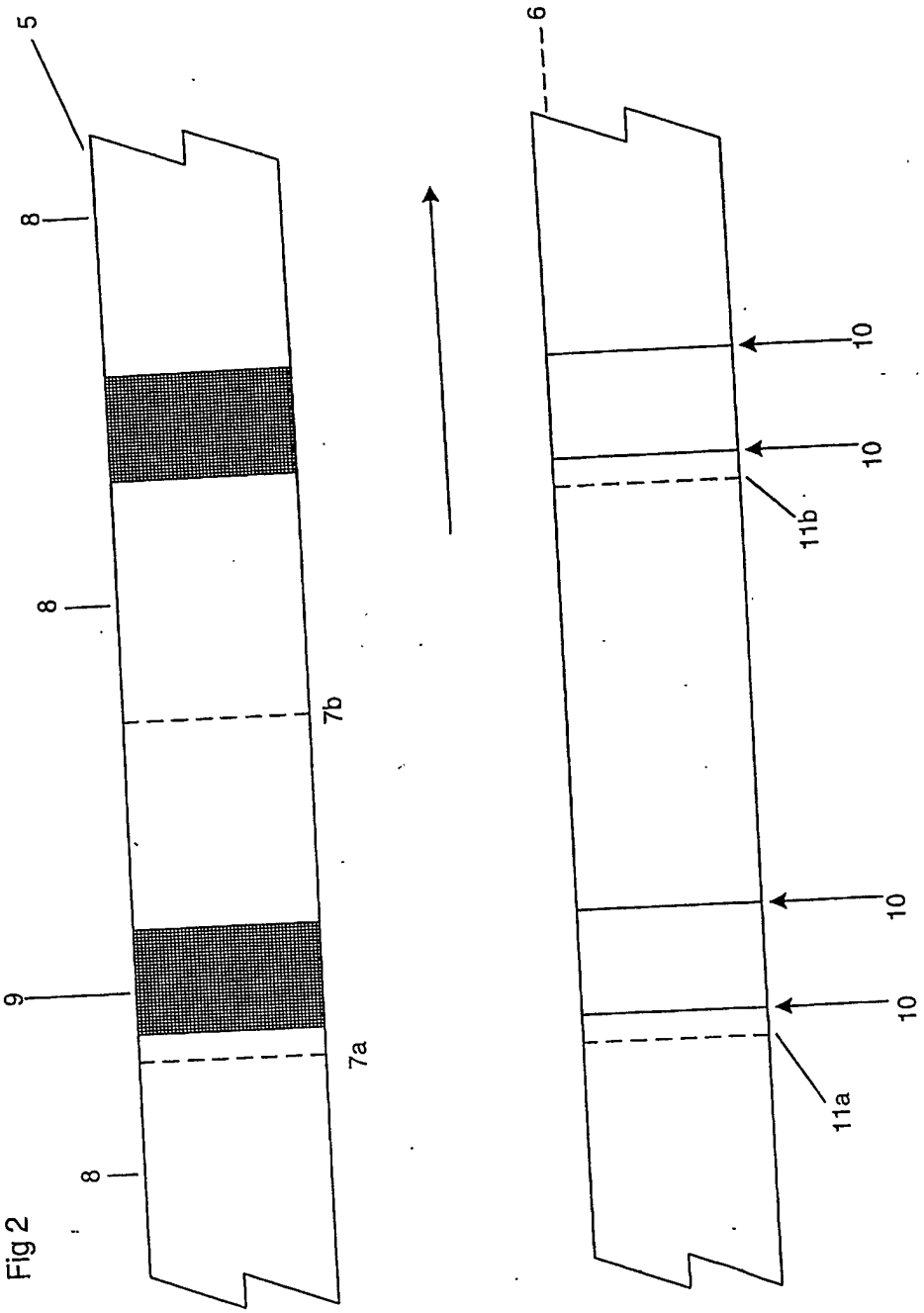


Fig 2

Fig 3a

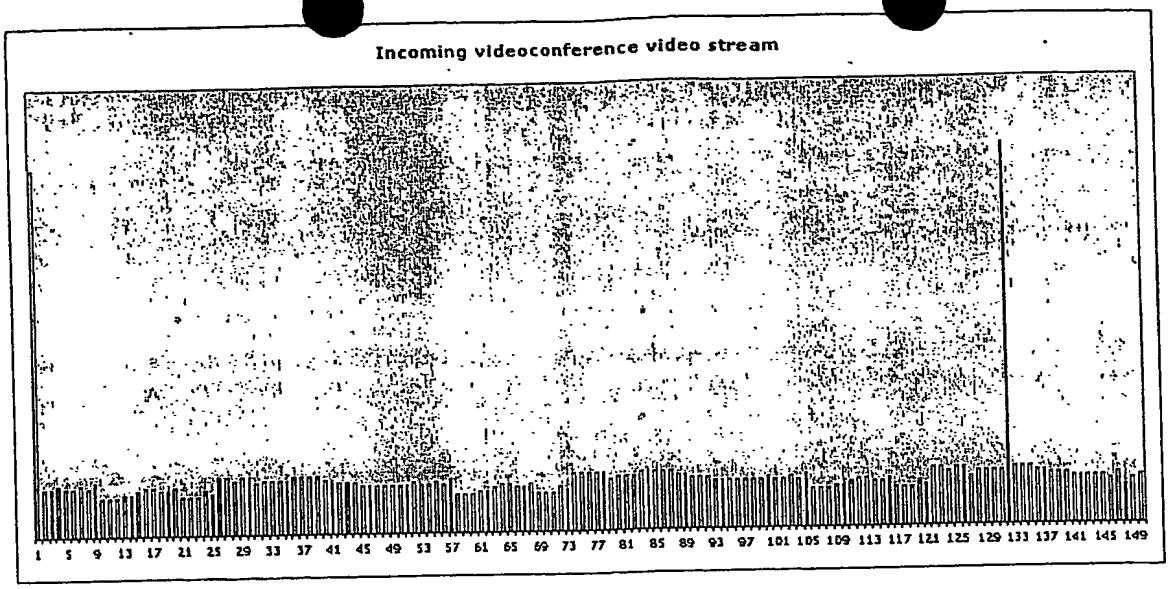


Fig 3b

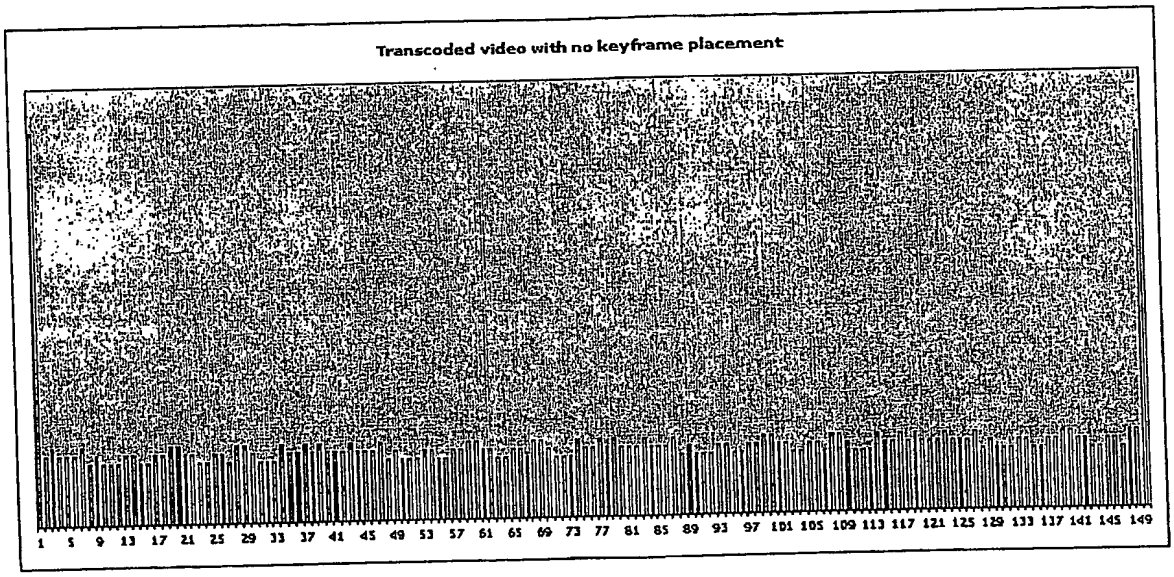


Fig 3c

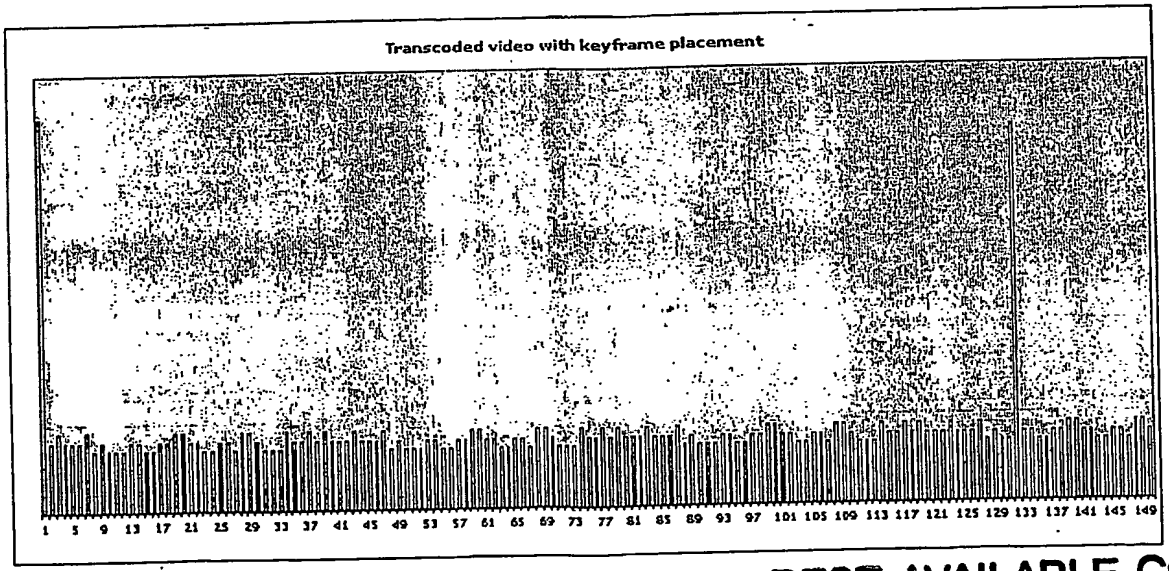


Fig 4

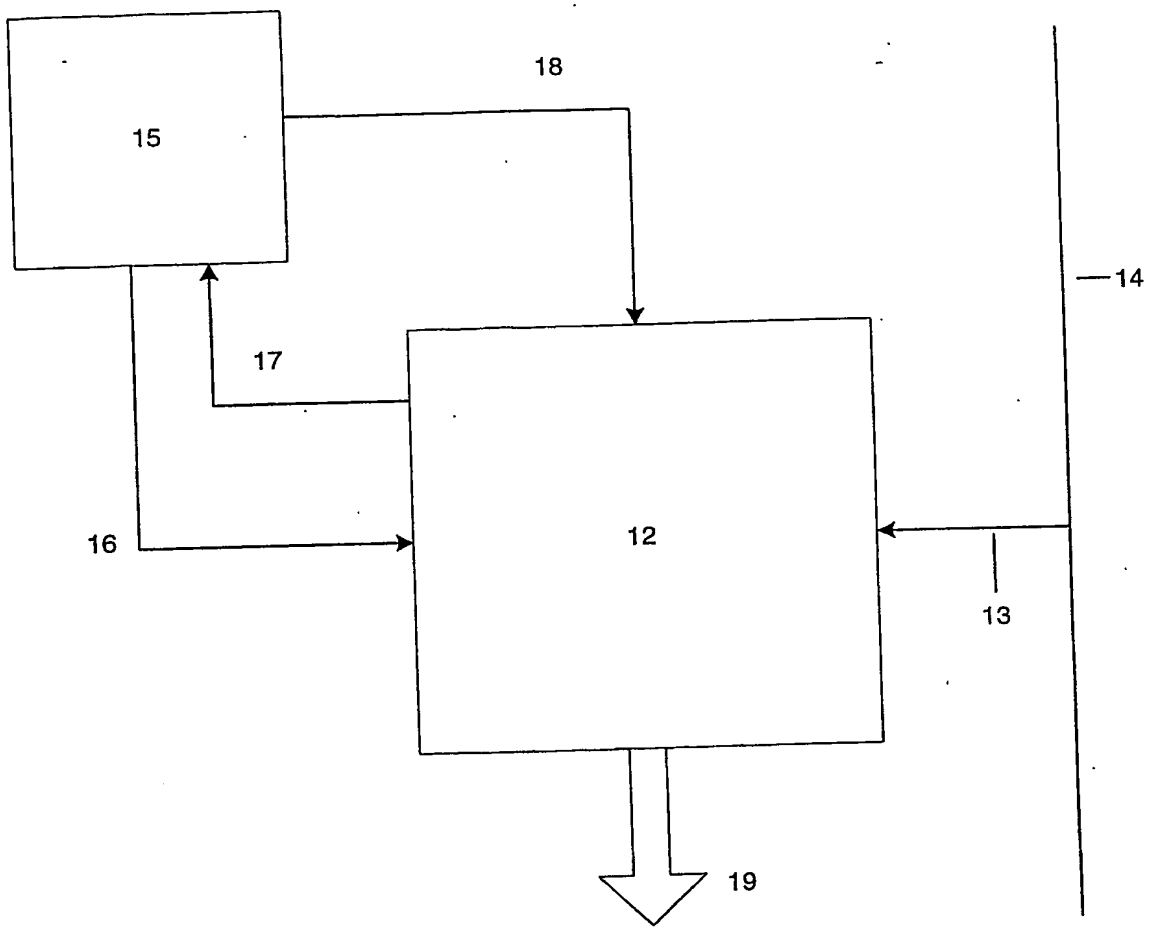


Fig 5a

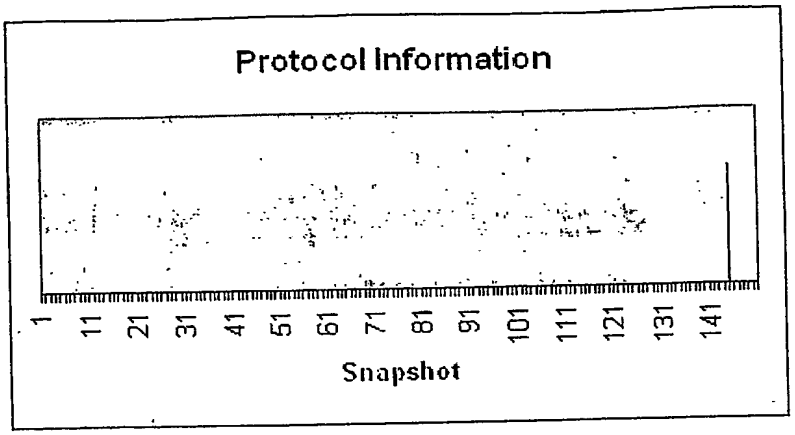


Fig 5b

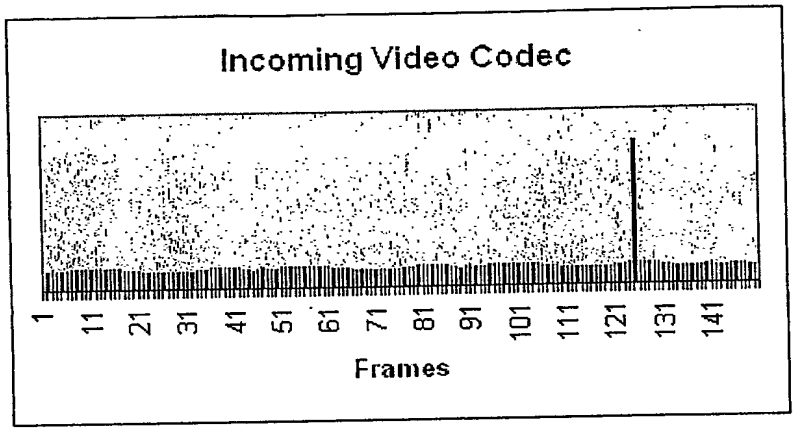


Fig 5c

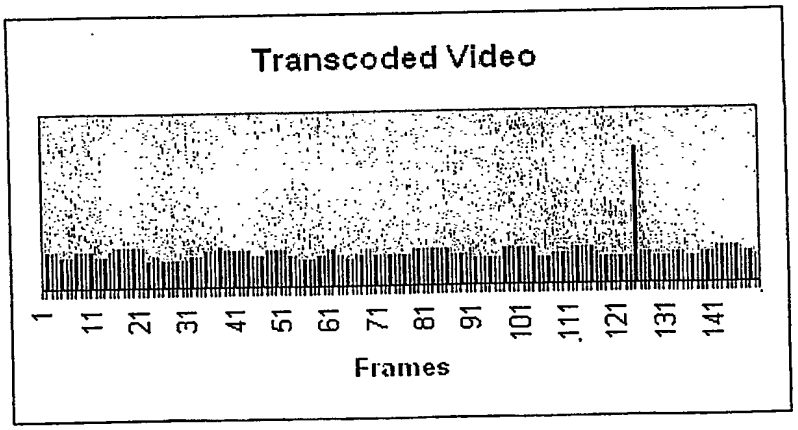


Fig 6a

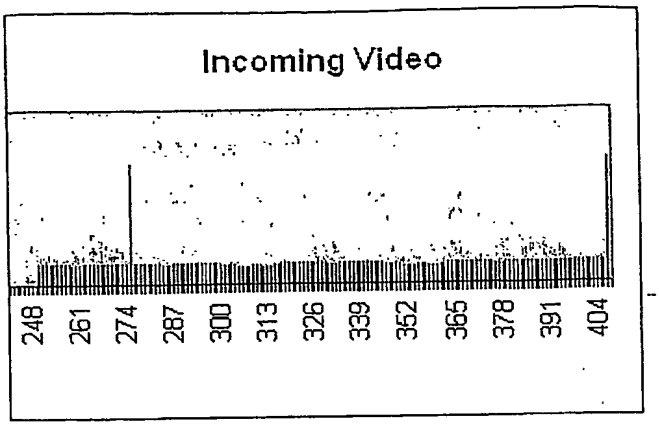


Fig 6b